

調音結合と個人性を考慮に入れた母音認識に関する研究

著者	西岡 公一
号	1011
発行年	1985
URL	http://hdl.handle.net/10097/9747

氏 名	にし	おか	こう	いち
	西	岡	公	一
授 与 学 位	工	学	博	士
学位授与年月日	昭和 60 年 10 月 9 日			
学位授与の根拠法規	学位規則第 5 条第 1 項			
研究科，専攻の名称	東北大学大学院工学研究科 (博士課程) 情報工学専攻			
学 位 論 文 題 目	調音結合と個人性を考慮に入れた母音認識に関する 研究			
指 導 教 官	東北大学教授 木村 正行			
論 文 審 査 委 員	東北大学教授 木村 正行	東北大学教授 城戸 健一		
	東北大学教授 曽根 敏夫	東北大学教授 丸岡 章		

論 文 内 容 要 旨

本研究は母音の自動認識のために重要な問題である調音結合および個人性の問題の解決を目指したものである。

しかし、調音結合と個人性の問題を同時に解決することは、きわめて困難と予想される。そこで、本研究では、最初に調音結合の問題と個人性の問題を分離して個々の問題の解決を計り、最後に両者の統合を計った。

論文は8章より構成されている。次に、各章の概要を述べる。

第1章は序論であり本研究の目的を述べている。

第2章では、最初に本研究で用いる音声資料・周波数分析法・特徴抽出法について述べた。ついで、従来の音声の自動認識において典型的に用いられてきた手法（従来法と呼び）を用いて認識実験を行い、従来法の問題点の検討および本研究の目標の明確化を計った。

この結果、単音節の母音の認識については、従来法では、すべての話者に対して一様に高い認識率を得ることが困難であることがわかり、話者適応の必要性が示された。

四連鎖母音に対する実験では、語中の母音の認識率が語頭・語尾の認識率より低い結果が得られ、調音結合への対策の必要性が示された。また、語中の母音の特徴パラメータの分布を作成して調音結合の影響について検討を行った。

文節中の母音に関する認識実験では、対称形の連鎖母音の方が非対称形連鎖母音より調音結合の

影響を受けやすいことが示された。

第3章は、1フレームの短時間スペクトルより得られる情報（以後、静的な情報と呼ぶ）による連鎖母音の認識について検討を行った。

はじめに、単音節の母音と4連鎖母音の語中の母音のスペクトルパターンの形状の比較（図1）を行った。

図より、語中母音のスペクトルパターンの形状は単音節のスペクトルパターンの形状から大きく変形しており、従来法では認識が困難であることが確認された。

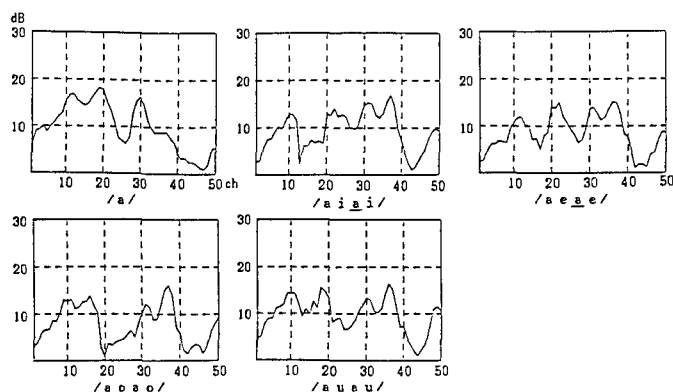


図1 単音節および連鎖母音中の /a/ の
スペクトルパターンの形状

しかし、両者の類似点について検討を行うと、第1ホルマントの周波数の位置については大きな変動がないことがわかった。

このことは、第1ホルマントの周波数を認識の際の特徴パラメータとして用いると調音結合を受けた母音でも正確な認識が可能になることを示唆している。ただし、ホルマントの抽出には不確実性がつきまとう。このため本研究では、低域のスペクトルを特徴として採用した。

次に、低域の形状の安定性について特徴パラメータの分布を作成し確認を行った。この結果、スペクトルパターンの低域の形状だけを用いた場合には、/i/と/u/および/e/と/o/の間の識別が困難になることが判明した。このため、低域の情報だけを用いて認識を行う際には、/a/と/i,u/と/e,o/の3母音間の認識を行うこととした。/i/と/u/および/e/と/o/間の識別は、スペクトルの高域もしくは全帯域の情報を用いて行うこととした。

認識実験を行った結果、一部の例外を除いて低域の特徴だけを用いた認識の有効性が示された。

従来の研究では、静的な情報を用いた連鎖母音の認識についての報告は見あたらない。

第4章では、連鎖母音中の特徴パラメータの時間的な動きを、認識系にどのように利用するかについて検討を行った。

最初に、連鎖母音中の特徴パラメータの動き（図2）について視察により検討を行った。この結果、ケプストラム係数の各次数において、ターゲットに接近している次数は、その時間的な変動が大きく、逆に時間的な変動の小さな次数については、その時間的な変動が小さいとの結論が得られた。このことから、認識においては時間的な変動の大きな次数については大きな重みづけを行い、逆に小さな変動の次数については小さな重みづけを行えば、正確な認識が可能になると考えられた。

そこで、ケプストラム係数の時間的な変動から認識時における各次数の重みづけを生成するモデルを考案した。次に、このモデルを用いて認識実験を行い、モデルの有効性を確認した。

従来の研究では、調音結合を受けることにより特徴パラメータの動きに平滑化が生ずると仮定し、特徴パラメータの動きを強調して、ターゲットに近づく試みが大半であった。

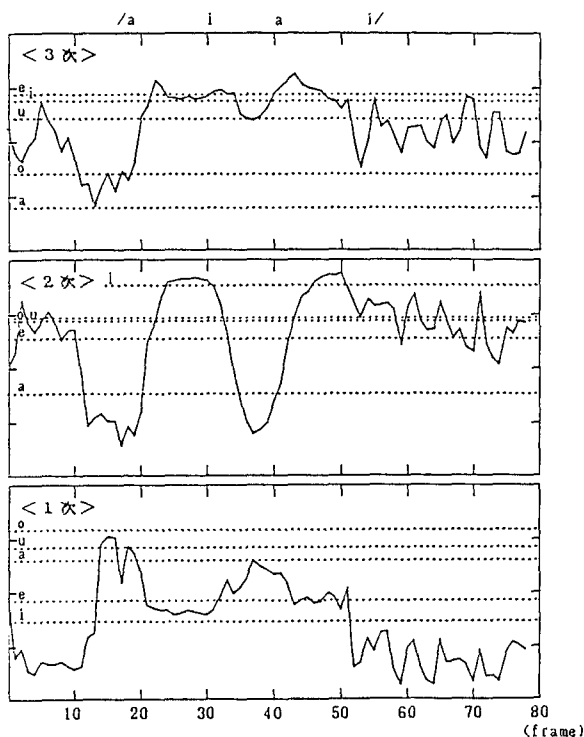


図2 四連鎖母音 /aiai/ における全帯域のケプストラム係数の振舞

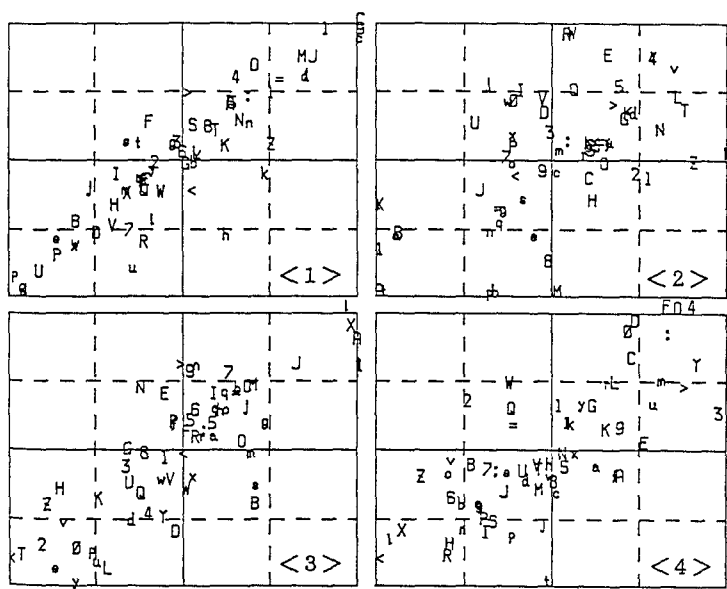


図4 各ケプストラム係数における話者ごとの重心と母音 /i/ の関係 (横軸-母音 縦軸-重心)

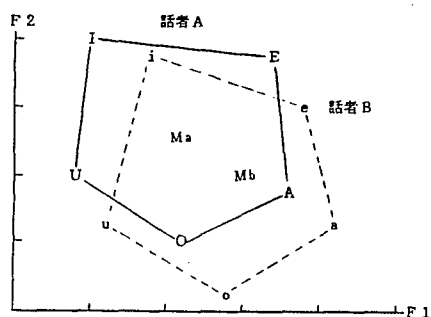


図3 五母音の重心による個人性のモデル化

これに対して本研究では、第2章と第4章より、ケプストラムを特徴に採用した場合には従来法が必ずしも最適でないことを示した。

第5章では、最初に個人性のモデル化を行い、次にモデルを話者適応に応用した。

個人性のモデル化は、話者の発声した五母音のケプストラム上での重心を用いて話者の個人性を表わした(図3)。

次に、多数の話者が発声した音声資料について各母音と重心の関係を調べた。この結果、重心と各母音の関係は多数の話者の間で線形な関係で記述し合えることが判明した。

さらに、重心と母音の関係を重回帰分析を用いて定式化を行った。

次に、上記の個人性のモデルを話者適応に応用した。単音節の母音の認識実験の結果、4個程度の学習母音で、第2章で行った従来法より高い認識率が得られた。認識結果を検討すると、90%以下の認識

率の話者が皆無であることや、モデルの定式化にあたって利用した係数設計用話者と非係数設計用話者間の認識率の差が、きわめて接近していることが判明した。これらのことは、重心を用いた個人性のモデルの有効性を示している。さらに、文節中の母音の認識に関して話者適応実験を行った。この結果、数個の文節による学習で適応が行われていることを確認した。

第6章では、重心を用いた個人性のモデルを話者識別に応用した。

話者識別においては、正確な個人性のモデル化が要求される。このため第6章は、より厳しい条件下のモデルの検証になる。

個人性のモデルに対するいくつかの改良後、66人に対する話者識別率は99%に達した。さらに、33人の話者の発声した非登録音声を用いた実験では、95%の識別率が得られた。

次に、話者登録時より6ヶ月もしくは1年経過してから発声された音声を用いて、長時間の経過に伴う特徴パラメータの変動に関する検討を行った。この結果、重心の位置も長時間の経過に伴う変動が認められた。このことから、重心だけでは個人性を表すには不十分であることが示された。

このため、時間の経過に対して安定な個人性のモデルについて検討を行った。

第7章は、第3章から第5章まで個別に検討を行い、その結果得られた手法の統合を計った。

最初に、低域の情報による認識と全帯域の情報による認識の統合を計った。2つの認識の統合とは、2つの手法により得られた認識結果より最適な認識結果を選び出すことである。ここでは、選択のための情報として、各候補母音の距離の時間変化パターンを採用した。この結果、より正確な認識が可能になった。

最後に話者適応を行い、全手法の統合を行い、調音結合と個人性の2つの問題の解決を目指した。

第8章は結論であり、本研究により得られた結果と今後の課題についてまとめた。

審 査 結 果 の 要 旨

音声の自動認識は、人工知能の研究の興味ある対象として、またマンマシンインタフェースの重要な手段として早くから研究され、話者数及び発声内容が強く制限された場合については、一部実用化されている。しかし、不特定話者や連続音声を対象とする場合には、個人差や調音結合などの物理的問題から言語レベルの問題まで、種々の困難な未解決の問題がある。

著者は、それらの問題の解決に寄与することを目的として、調音結合と個人性を考慮した母音認識について研究した。本論文はその成果をまとめたもので全編8章よりなる。

第1章は序論である。

第2章では、66名の話者による単音節母音、発声の速さを変えた四連鎖母音、及び文節中の母音からなる資料について、まず従来の統計的認識手法を用い調音結合や個人差が認識結果に与える影響を詳細に分析し、本研究で扱う問題の性格を明確にしている。

第3章では、調音結合に殆ど影響されない特徴量が第1ホルマントを中心とする低域に存在することを見出し、この領域のケプストラム係数でその特徴量を代表させることにより、それを母音認識に活用する有用な方法を与えている。

第4章では、連鎖母音の1次から10次までのケプストラム係数の時間的変化を分析し、調音結合により時間的変化が平滑化される係数の重みを自動的に小さくする動的重み付けにより、連鎖母音に対しても高い認識率を得る方法を示している。

第5章では、多数話者の場合に特徴空間における5母音の分布に重なりが生じるが、特定話者では良好に分離することに着目し、ケプストラム係数空間での5母音の重心と各母音の位置との間の関係式を個人性のモデルとして導入し、発声された母音から5母音の重心と他の母音の位置を推定して話者適応を行う方法を与え、認識実験によりその有用性を立証している。第6章では、個人性のモデルが話者認識にも有用であることを示している。

第7章では、第3章と第4章の方法を相互補完的に用いて、調音結合に対しても有効な母音認識の方法を与えるとともに、この方法と第5章の話者適応の方法とを結合することにより、不特定話者に対しても有効な母音認識の手法を得ている。第8章は結論である。

以上要するに本論文は、調音結合の影響が小さい特徴量を活用する方法と個人性のモデルに基づく話者適応の方法とを統合することにより、調音結合と不特定話者に対しても有効な母音認識の一手法を開発し、音声認識に新知見を加えたもので、情報工学の発展に寄与するところが少なくない。

よって、本論文は工学博士の学位論文として合格と認める。